

Using Deep Learning Based Approaches for Bearing Fault Diagnosis with AE Sensors

Miao He¹, David He¹, and Eric Bechhoefer²

¹*Department of Mechanical and Industrial Engineering, University of Illinois at Chicago, Chicago, IL, 60607, U.S.*

*mhe21@uic.edu
davidhe@uic.edu*

²*Green Power Monitoring Systems, Cornwall, VT, 05753, U.S.*

eric@gpms-vt.com

ABSTRACT

In the age of Internet of Things and Industrial 4.0, the prognostic and health management (PHM) systems are used to collect massive real-time data from mechanical equipment. Mechanical big data has the characteristics of large-volume, diversity and high-velocity. Effectively mining features from such data and accurately identifying the machinery health conditions with new advanced methods become new issues in PHM. A major problem of using the existing PHM methods for machinery fault diagnosis with big data is that the features are manually extracted relying on much prior knowledge about signal processing techniques and diagnostic expertise, limiting their capability in fault diagnosis with big data.

This paper presents a deep learning based approach for bearing fault diagnosis using acoustic emission (AE) sensors with big data. Different from widely used shallow neural network architecture with only one hidden layer, the branch of machine learning methods with multiple hidden layers are regarded as deep learning method. The presented approach pre-processes AE signals using short time Fourier transform (STFT) and extract features. Based on the simply processed AE features, an optimized deep learning structure, large memory storage retrieval neural network (LAMSTAR) is used to perform bearing fault diagnosis. The unique structure of LAMSTAR enables it to establish more efficient and sparse distributed feature maps than traditional neural networks. By leveraging the labelled information via supervised learning, the trained network is endowed with discriminative ability to classify bearing faults. The AE signals acquired from a bearing test rig are used to validate the presented method. The test results show the accurate

classification performance on various fault types under different working conditions, namely input shaft rotating speeds. It also proves to be effective on diagnosing bearing faults with relatively low rotating speeds.

1. INTRODUCTION

Bearings are widely used as vital components in modern mechanical equipment. The occurrence of bearing faults will result in significant breakdown time, elevated repair cost and even a potential decrease in productivity. In the age of Internet of Things and Industrial 4.0, the prognostic and health management (PHM) systems are used to collect massive real-time data from mechanical equipment. Mechanical big data has the characteristics of large-volume, diversity and high-velocity. Effectively mining features from such data and accurately identifying the machinery health conditions with new advanced methods become new issues in PHM.

Over the recent years, the application of acoustic emission (AE) sensors for machine health monitoring and fault diagnosis has been reported (Morhain & Mba, 2003, Mba, 2008, He *et al.*, 2011, He & Zhang, 2012, Nienhaus *et al.*, 2012, Hemmati *et al.*, 2016,). In comparison with other sensor signals such as vibration signals, AE signals have certain advantages in capturing and representing both local and global bearings fault features. A key step in using AE sensors for machine fault diagnosis is signal processing and valuable feature extraction. In machine health monitoring and fault diagnosis, the AE signals are used to detect, locate and characterize damage (Girado *et al.*, 2003, Nigam *et al.*, 2004). Mba (2008) reported the efficiency of fault detection and health monitoring by using Hilbert-Huang transform (HHT) on rotational machinery system and components. Along with application of HHT, another threshold based technique was used to increase the signal-to-noise ratio of targeted AE burst signal (Mba, 2008). Hemmati *et al.* (2016) proposed a new algorithm by using wavelet packet

Miao He *et al.* This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

transform to optimize the ratio of kurtosis and Shannon entropy for diagnosing local defects on bearings under various working conditions. The validation result showed effectiveness on extracting bearing characteristic frequencies with background noise under various operation conditions. In a recent study (He *et al.*, 2011), a data mining based full ceramic bearing fault diagnostic algorithm using AE sensors was presented. Condition indicators were extracted from decomposed intrinsic mode functions components of the AE signals using HHT. The extracted condition indicators were then used in both k -nearest neighbor (KNN) and back propagation (BP) neural network for bearing fault diagnosis.

A major problem of using the existing PHM methods for machinery fault diagnosis with big data is that the features are manually extracted relying on much prior knowledge about signal processing techniques and diagnostic expertise, limiting their capability in fault diagnosis.

Different from widely used shallow neural network architecture with only one hidden layer, the branch of machine learning methods with multiple hidden layers are regarded as deep learning method. Deep learning attempts to model complexity and internal correlation in dataset by using multiple processing layers, or with complex structures, to mine the information hidden in dataset for classification or other goals (Hinton & Salakhutdinov, 2006). As a deep learning method, LAMSTAR has been applied in multiple fields such as image recognition (Girado *et al.*, 2003), biomedical diagnosis (Nigam *et al.*, 2004, Sivaramakrishna & Graupe, 2004, Waxman *et al.*, 2010, Isola *et al.*, 2012), with solid result showing the ability of LAMSTAR for rapidly processing large amount of data and less error percentage than regular machine learning algorithms. As one of the representative classical neural networks, the BP neural networks have been widely used in machinery fault diagnosis (Paya *et al.*, 1997, Huang *et al.*, 2007). However, the reported studies based on BP neural networks have shown the strict requirement of signal pre-processing using complicated signal processing methods including HHT and wavelet packet transformation. With the introduction of forgetting, rewarding/punishing features into a traditional neural network, LAMSTAR works on a closer level of mimicking the working process of a natural brain. Without the limitation on network size, LAMSTAR can grow/shrink in dimension without changing the original structure and maintain fast training speed. As an input, the LAMSTAR network accepts data defined by the user, such as system state and system parameter. Then, the system builds a model

to be trained for memorizing the information of input data and correlation between input data. The trained module can be used in testing, by searching the stored knowledge to find the best approximation to the features of input data. The standard perceptron-like neurons are employed in LAMSTAR, arranging in self-organizing maps (SOM) modules. The SOM structures in LAMSTAR are governed by winner taking all (WTA) strategy and the memories of those neurons in SOM are stored in bidirectional associative memory (BAM) function. The optimization of a LAMSTAR network is to determine the link weights that store relation message between various SOM modules and between neurons in SOM and output decision layer. The link weights contribute along with information stored in winner neuron from each individual SOM to SOM-type output decision layer. With all of the characteristics discussed above, a LAMSTAR network can learn and understand system information more systemically and intelligently. Yoon *et al.* (2013) have successfully applied LAMSTAR on AE signals for full ceramic bearing fault diagnosis. AE signals were pre-processed by using HHT to extract conditional indicators as input to BP, LAMSTAR, and KNN. The classification result from that study shows faster learning speed and higher accuracy from LAMSTAR. Deep learning represents an attractive option to process mechanical big data for machinery fault diagnosis.

In this paper, a LAMSTAR based bearing fault diagnosis using AE sensors is presented. The unique contribution of the paper is that the AE are pre-processed with the simple short-term-Fourier-transform (STFT) method rather than computationally complicated signal processing algorithms such as wavelet transform and HHT. The AE features obtained using STFT are directly taken as inputs to LAMSTAR for bearing classification. Thus, the feature extraction process for fault diagnosis is greatly simplified.

2. THE METHODOLOGY

The general procedure of applying LAMSTAR network for AE sensor based bearing fault diagnosis is shown in Figure 1. As shown in Figure 1, the AE data is first pre-processed using STFT to generate a spectrum matrix S . Sub-patterns are then generated from the spectrum matrix S and used to obtain the optimal LAMSTAR network model for bearing fault diagnosis. The basic LAMSTAR network structure and specific design of LAMSTAR network in this paper are explained next.

2.1. Basic Structure of LAMSTAR Networks

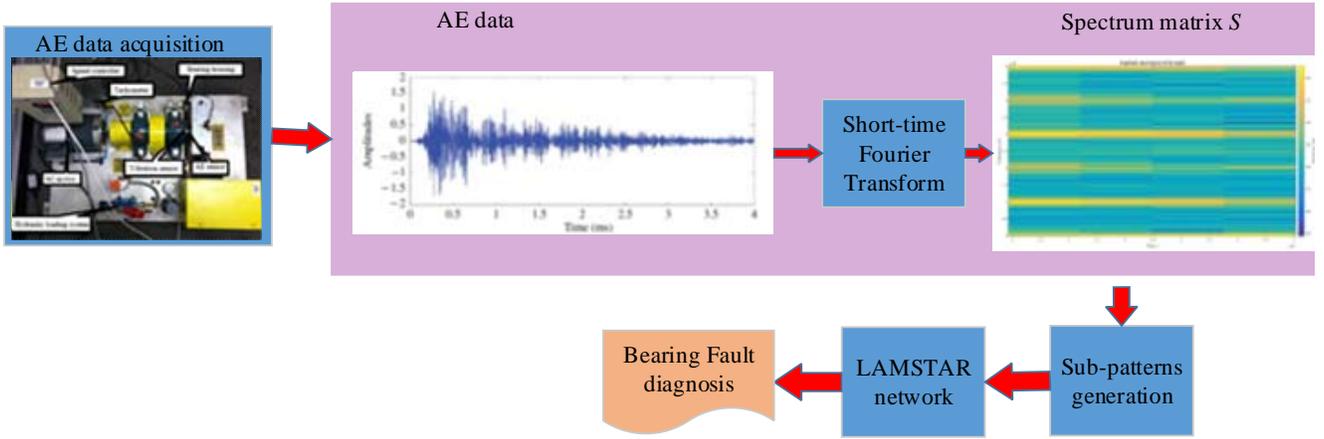


Figure 1. LAMSTAR for AE based bearing fault diagnosis

The LAMSTAR network is specifically designed for retrieval, classification, prediction and decision making problems, especially those involving large amount of categories. With advantages of SOM-based network modules and statistical decision tools, LAMSTAR is capable of pattern storage and retrieval. The information in network is continuously being updated for each sample through learning and correlation, to make LAMSTAR an intelligent system. One unique function of LAMSTAR is to handle analytical and non-analytical data, with complete or some missing categories information. Such uniqueness is achieved by implying forgetting, interpolation and extrapolation features, allowing the network to reduce the weights of stored information and still being able to approximate forgotten information by extrapolation or interpolation.

The decision making principle in LAMSTAR is the same as the classical neural networks. If n denotes the number of inputs fed into j^{th} neuron as $\{v_{ij}, i = 1, 2, \dots, n\}$, then output y_j of the j^{th} neuron can be expressed as:

$$y_j = f_N(\sum_{i=1}^n w_{ij} v_{ij}) \quad (1)$$

where $f_N(\cdot)$ represents nonlinear activation function. Variable w_{ij} are the weights assigned to the i^{th} inputs of j^{th} neuron and whose setting is the learning action of the LAMSTAR network. The information in LAMSTAR is stored and processed via correlation links between individual neurons in separate SOM modules. Given a coded real matrix X as input pattern:

$$X = [x_1^T, x_2^T, x_3^T, \dots, x_N^T] \quad (2)$$

where x_i^T stands for transpose of sub-pattern x_i . Each sub-pattern x_i is channeled to a corresponding i^{th} SOM module that stores data focusing on i^{th} category of the input pattern. A general structure of LAMSTAR network is presented in

Figure 2. In Figure 2, input pattern represents each signal to be diagnosed, containing all of sub-patterns generated from the spectrum matrix S . Input modules represent all of the structure before final decision module, including input pattern and SOMs. Considering each input pattern as an input layer, each sub-patterns can be viewed as an input neuron and each SOM module as a hidden layer. The LAMSTAR network does not create neurons for an input pattern. Instead, only individual sub-patterns are stored in SOM layers, and correlations between sub-patterns are stored as link weights (Graupe & Kordylewski, 1997). The sub-pattern construction is explained in details in Section 2.2.1. How a link weight in LAMSTAR is created and adjusted is explained next.

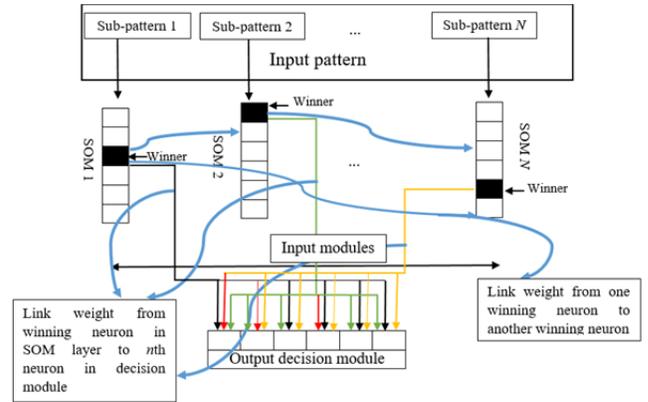


Figure 2. General structure of the LAMSTAR network

When a new input pattern is loaded into a LAMSTAR network, the LAMSTAR network checks all the storage-weight w_i in the i^{th} SOM module corresponding to the i^{th} sub-pattern to be stored. If any existing sub-pattern matches the input sub-pattern x_i within a pre-defined error tolerance, then it is claimed as the winning neuron for that particularly

observed input sub-pattern. For most applications with storage of purely numerical input sub-patterns, the storage of such sub-patterns into SOM module can be simplified by directly mapping each SOM module into a pre-defined range of value. For example, the i^{th} sub-pattern with value of 0.6 will be stored in the neuron of the i^{th} input SOM module representing range of 0.5 to 0.75. The searching procedure and decision of winning neuron in each module for SOM dynamic weights construction is explained in Section 2.2.3.

The correlation between sub-patterns is stored in link weights that connect neurons in different SOM modules. Also, the correlation between winning neurons in each input SOM module and decision SOM module are stored as link weights. Link weights are functioned as links among knowledge learned by brain. Thus, the link weights become fundamental to allow interpolation and exploration of sub-patterns. The link weights are updated as for a given input pattern with a determined winning k^{th} neuron in i^{th} SOM module and a winning m^{th} neuron in j^{th} SOM module. A winning neuron is determined for each input sub-pattern based on the similarity between the input sub-pattern and a weight vector \mathbf{w} (stored information). For a given sub-pattern \mathbf{x}_i , the winning neuron is determined by minimizing the distance norm $\|\cdot\|$ as below:

$$d(i, j) = \|\mathbf{x}_i - \mathbf{w}_j\| \leq \|\mathbf{x}_i - \mathbf{w}_k\|, \forall k \neq j \quad (3)$$

The vector \mathbf{x}_i will be stored in weights w_{ij} of vector \mathbf{w}_j relating to the j^{th} neuron when the distance satisfies Eq. (3). Vector \mathbf{w}_j represents the storage weight vector only for the input SOM modules. Decision SOM module does not store the information of input patterns. The link weight $L_{i,j}^{k,m}$ is calculated by adding a pre-defined reward. Meanwhile, all other links can be decreased with a predefined punishment. The link weights update can be expressed as followings:

$$L_{i,j}^{k,m}(t+1) = L_{i,j}^{k,m}(t) + \Delta R \quad (4)$$

$$L_{i,j}^{k,m}(t+1) = L_{i,j}^{k,m}(t) - \Delta P \quad (5)$$

$$L(0) = 0 \quad (6)$$

where, in Eq. (4) and Eq. (5), $L_{i,j}^{k,m}$ donates links between winning neuron i in k^{th} module and winning neuron j in m^{th} module, ΔR and ΔP are pre-defined reward and punishment values, and t represents the number of the iterations that the link weights are updated. The initial link weight is set as 0. The value of reward and punishment are normally pre-defined between 0 and 0.5, and they can be set as same value. The output that matches with target input will be rewarded by a non-zero increment; otherwise it will be punished by a non-zero decrement. It is computationally efficient in most applications that only link weights between winning neurons in the input SOM modules and decision SOM module are updated.

The output at the decision SOM modules is made by analyzing correlation links between decision neurons in the

decision SOM module and neurons in all input SOM modules. To make such a decision, LAMSTAR produces a winning decision neuron n from the set of output neurons J in the decision SOM module by searching for the neuron in output module with the highest cumulative value of link weights connecting to the selected winning neurons in each input modules. The equations to make such a decision for the i^{th} output SOM module are given as follows:

$$E(j) = \sum_{\kappa w^*}^M L_{\kappa w^*}^{i,j}, \forall j \in J \quad (7)$$

$$E(n) \geq E(j), \forall j \in J \quad (8)$$

where i donates the i^{th} output module; n donates winning neuron in the i^{th} output module; κw^* donates winning neuron in every κ^{th} input module; M stands for the total number of input modules; $L_{\kappa w^*}^{i,j}$ stands for link weight between winning neuron in κ^{th} input module and neuron j in i^{th} output module; and J is set of the neurons in i^{th} output module. Thus, $E(j)$ represents the sum of link weights connecting to the j^{th} neuron in i^{th} output SOM module from winning neuron κw^* in every κ^{th} input module, and $E(n)$ represents the sum of link weights connecting to the winning neuron n in i^{th} output SOM from winning neuron κw^* in every κ^{th} input module.

In the LAMSTAR based bearing fault diagnosis approach, the algorithm will first use the training data to update the link weights. More specifically, it will find the winning neuron in each layer and applies the WTA principle to update the link weights. In testing procedure, the LAMSTAR network will calculate the winning neuron in the decision layer serving as the label information. In the last step, the LAMSTAR will provide the accuracy rate defined as ratio of success classification number against total number of tested data points.

2.2. The Design of the LAMSTAR Network for AE based Bearing Fault Diagnosis

The design of the LAMSTAR network for AE based bearing fault diagnosis involves the following tasks: sub-pattern generation, input data normalization, dynamic formation of neurons in SOM, determination of the link weights, and neural network test. They are explained next.

2.2.1. Sub-pattern generation

The basic storage modules of the LAMSTAR networks are SOM modules as discussed in Section 2.1. In LAMSTAR networks, the information is stored and processed via correlation links between individual neurons in separate SOM modules. The link weights are the main engine of the network, connecting every SOM module so that the emphasis is on correlation of link weights between atoms of memory, not on the memory atoms themselves. In such case, the input data should be modeled as sub-patterns for each

SOM module to store and process, and each sub-pattern describes characteristics or attributes of the input pattern.

The collected raw AE data will be pre-processed by STFT at first. Then, the processed data are transferred into spectrum 2-D matrix with a size of $l \times n$. Given an AE signal $s(t)$, the spectrum matrix can be written as $S_{l \times n}$ with a size of $l \times n$, where l denotes row number and n denotes column number. Each of the AE signal is viewed as one pattern in this case. By considering the spectrum matrix as a spectrum plot with known elements inside, the sub-sampling method used in LAMSTAR based image recognition application (Girado, 2004, Homayon, 2015) can be applied to generate sub-patterns. Sub-patterns are generated by taking samples from previously generated spectrum matrix, by using a sliding box.

For each pattern, data subsets are sampled using sliding box with a size of $d \times d$ by sliding the spectrum matrix from left to right, then top to bottom, sequentially. Every subset obtained from sliding box will be transformed column by column into a 1-D vector, taken as one sub-pattern in a SOM. The procedure for obtaining sub-pattern from spectrum matrix is presented in Figure 3.

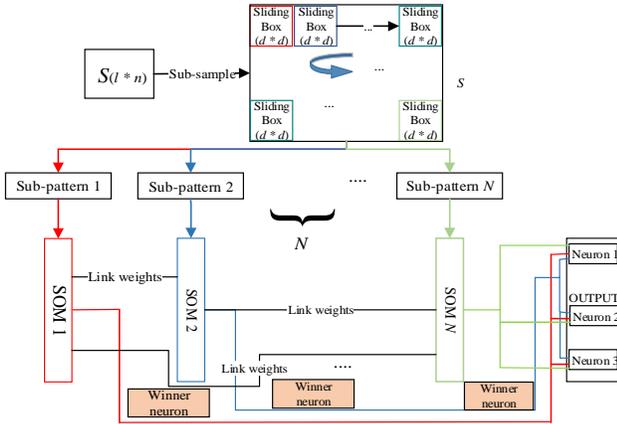


Figure 3. The procedure for obtaining each sub-pattern from spectrum matrix as input to LAMSTAR

With a defined size of sliding box as $B_{d \times d}$, the total number of generated sub-patterns N is decided as:

$$N = (l/d) \times (n/d) \quad (9)$$

Normally the value of N is set as an integer by adjusting the value of variable d in an acceptable range. And for each sub-pattern, the number of elements is d^2 . Since each spectrum matrix $S_{l \times n}$ is considered as one pattern, then one can define a sub-pattern as Ω_v , $v = 1, 2, \dots, N$. Thus a sub-pattern can be written as:

$$\Omega_v(f, g) = S_{l \times n} \left(\left\lfloor \frac{N}{n} \right\rfloor \times f, \text{mod} \left(\frac{l}{d} \right) \times g \right), \quad \forall f, g < d \quad (10)$$

The data sampled from the spectrum matrix is stored in sub-patterns, and each sub-pattern will be modeled as one SOM layer. Thus, the number of SOM is the same as the number of generated sub-patterns, as related with spectrum matrix dimension and sliding box size directly. A standard procedure for selecting the dimension of sliding box has not been reported. Currently, the sliding box dimension is selected by trial and error approach. The goal of determining an appropriate sliding box dimension is to select a sub-pattern containing as many variant information and thus can be used as a feature. Therefore, the entropy of elements per sliding box can be used as one criterion when selecting the appropriate sliding box dimension. In bearing fault diagnosis, each fault will be represented by an output neuron firing sequence on a decision layer. Therefore, the decision layer should contain enough number of output neurons such that a complete permutation of the output neuron firing sequences can be used to represent the patterns (i.e., the bearing conditions) to be classified.

2.2.2. Input normalization

For each sub-pattern $x_i = [x_1, x_2, \dots, x_j, \dots]$, the normalization of x_i is computed as:

$$x_i^{norm} = \frac{x_i}{\sqrt{\sum_j x_j^2}} = \left[\frac{x_1}{\sqrt{\sum_j x_j^2}}, \frac{x_2}{\sqrt{\sum_j x_j^2}}, \dots, \frac{x_j}{\sqrt{\sum_j x_j^2}}, \dots \right] \quad (11)$$

Thus, the input sub-pattern is normalized between 0 and 1.

2.2.3. Dynamic formation of neurons and weights in SOM

In this paper, the SOM models are built dynamically instead of setting a fixed number of neurons arbitrarily. The network is built to have neurons depending on the class to which a given input to a particular sub-pattern might belong. Such designed network produces less number of neurons and the time required to fire a particular neuron at the classification stage is reduced considerably.

The first neurons in all the SOM modules are constructed as Kohonen neurons and are placed as the first ones in each list of the neurons. One neuron is built with inputs and randomly initialized weights to start with and they are normalized following the same equation for input sub-patterns. Then the weights are updated until the output of the neuron is made equal to 1 with pre-defined error tolerance. Let $w(n)$ and $w(n+1)$ be the weight at iteration n and $n+1$, then the weight is updated as:

$$w(n+1) = w(n) + \alpha \cdot [x - w(n)] \quad (12)$$

where learning constant α is set as 0.8 and x denotes sub-patterns.

The output value of neuron is computed as:

$$z = \mathbf{w} * \mathbf{x}^T \quad (13)$$

In the dynamic SOM weight construction process, all of incoming sub-patterns are checked if they are zero sub-patterns (i.e. zero vector). The trained output weight is set to be zero and searching winning neuron step is skipped if zero sub-pattern exists. Otherwise, any incoming sub-pattern searches among the previously constructed neurons and corresponding weights if any neuron generates output z that equals to 1 with pre-defined error tolerance. The neuron satisfies with output that equals to 1 with pre-defined error tolerance is claimed as winning neuron. If the searching of winning neuron fails, another neuron and corresponding weight set are constructed additionally with pre-defined error tolerance.

After all of the sub-patterns are imported into the respective SOM modules the output at any of the previously built neuron is compared to 1 with pre-defined error tolerance. The neuron who satisfies the condition will be rewarded with a non-zero increment and punished with a small non-zero decrement.

2.2.4. Determination of the link weights

Link weights store the correlation between sub-pattern in different input SOM modules, and correlation between the winning neurons in different input SOM module and neurons in decision SOM module. The value of link weight will be changed in every iteration according to reward/punish policy. In a modification version of LAMSTAR, the link weight $L_{i,j}(m,k)$ from neuron m in SOM module to neuron j in the i^{th} decision layer is replaced by a normalized link weight as:

$$L_{i,j}^{\text{norm}}(m,k) = L_{i,j}(m,k)/n(m,k) \quad (14)$$

where $n(m,k)$ denotes the number of times that neuron m is the winning neuron in k^{th} SOM module.

2.2.5. The test of the LAMSTAR neural network

Testing the network follows a straightforward approach. The test samples are preprocessed and normalized into sub-patterns the same way as the training set. The stored weights of the SOM module and the link weights, obtained during training of the network, are used. The input is then propagated through the network and output is determined. A slightly different code is used for testing the network where the weights of the SOM and link weights are loaded from the already stored values obtained from the training process.

3. EXPERIMENT SETUP AND LAMSTAR NETWORK DESIGN

3.1. Bearing Seeded Fault Test Experiment Setup

This section covers the experimental setup used to validate the deep learning based AE bearing fault diagnostic technique. Figure 4 shows the bearing test rig used to collect the AE data and conduct the bearing seeded fault tests. A wide band (WD) type AE sensor was axially mounted on the face of the bearing housing using instant glue.

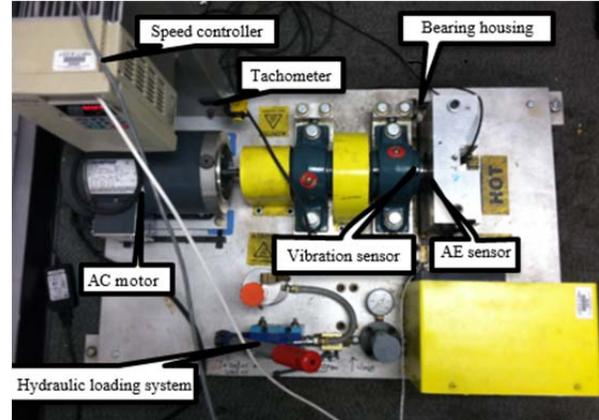


Figure 4. The bearing test rig

Type 6205-2RS steel FAG ball bearings were used for testing. Four fault types were simulated on steel bearings: inner and outer race faults, rolling element fault, and cage fault (see Figure 5). The inner and outer race faults were generated by scratching the steel race surfaces with a diamond tip grinding wheel bit to cover the ball contact surface. The ball fault damage was created by cutting the steel cage in one of the ball locations and then using the diamond tip grinding wheel bit to create a small dent in one of the steel balls. For the cage fault, the steel cage was cut in between two ball locations. For all seeded fault tests, the bearing seal and grease was removed and replaced following the creation of the fault.



Figure 5. The bearing seeded faults

Figure 6 shows the AE data acquisition system consisting of a demodulation board, power supply, along with the function generator and sampling device. The demodulation board performs the multiplication of the AE sensor signal with the reference signal output from the function generator which allows a sampling frequency reduction technique to

be implemented. It takes the two signals as inputs and the output is the multiplication of the signal inputs. After shifting the signal information to a lower frequency range, the output is fed to the sampling board while filtering out the high frequency component. To down shift the AE signal frequency, its carrier frequency had to be determined in order to introduce a reference signal for demodulation. Thus, the goal was to determine the central AE carrier frequency and set the reference signal frequency as close to it as possible.

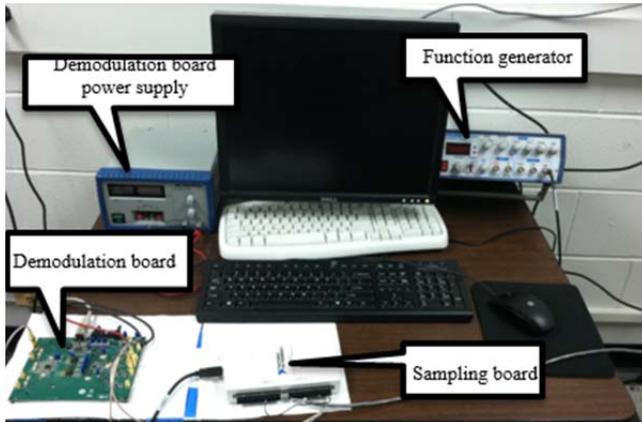


Figure 6. The AE data acquisition system

A sweep function, created by the function generator, was used to record the output of the system. After examining the energy envelope at different frequencies, an estimate of the frequency range of the output was found. It was determined that the central AE signal carrier frequency was 400 kHz and was used as the demodulation reference frequency. The sampling device is a low frequency data acquisition device capable of handling a sampling frequency up to 250kS/s. NI Labview signal express was used for data acquisition, collecting the continuous AE signals at a sampling rate of 100 kHz. Both the healthy bearing and the seeded fault bearings were tested at following shaft speeds: 2 Hz, 4 Hz, 6 Hz, 8 Hz, 10Hz, 30 Hz, and 45 Hz. At each speed, 5 samples were collected for a total of 15 samples of AE data per bearing type. For consistency, the AE sensor was placed in the same axial location for all data acquisitions.

3.2. The LAMSTAR Neural Network for the Validation Test

Each of the AE signal dataset was transformed into a 2-D spectrum matrix as explained in Section 2.1. 10 patterns from each of the 5 bearing conditions: inner race fault, outer race fault, cage fault, ball fault and healthy condition were generated. Therefore, there were a total of 50 patterns generated.

In the implementation of the LAMSTAR network, each spectrum matrix representing one pattern has a dimension of 250×4000 . A sliding box with size of 50×50 was selected

initially. Thus, 40 SOMs (hidden layers) were generated for 40 sub-patterns sampled from each spectrum matrix. In each SOM module, the dynamic neurons were constructed for storing representative value of cell in matrix. Thus, the number of neurons in each SOM varied from 0 to 2500. Considering there are 5 bearing conditions in this study: inner race fault, outer race fault, cage fault, ball fault and healthy condition, three LAMSTAR network output neurons were used to give a complete permutation of 6 firing sequences with each sequence representing a condition. Table 1 shows the bearing conditions and their LAMSTAR output neuron firing sequence representations. The error tolerance was set to be 10^{-9} ; and learning rate alpha was set to be 0.8 as a constant, and error tolerance for winning neuron decision in dynamic SOM weights construction was set to be 10^{-7} . Based on the principle of ANN application, the collected dataset was divided into training and validation groups for training and validating generated network. In this paper, the ratio of training to testing dataset was 60% to 40% against total. The graphic procedure of LAMSTAR network structure in this paper is displayed in Figure 2 as above.

Table 1. Bearing conditions and their LAMSTAR output neuron firing sequence representations

Bearing condition	Output neuron 1	Output neuron 2	Output neuron 3
Inner race fault	0	0	1
Outer race fault	0	1	0
Cage fault	1	0	0
Ball fault	1	0	1
Healthy	1	1	0

4. THE RESULTS

Using the AE signals collected during the bearing seeded fault tests, 50 patterns were used to train the LAMSTAR model, 10 for each bearing condition. The dynamically built neurons in SOM modules enable the large reduction on training time as the search time to find the winning neuron was reduced to a small number of neurons in many cases. The neural network learns as it goes even if untrained. In addition to LAMSTAR, another deep learning algorithm convolutional neural network (CNN) was used to perform the bearing fault diagnosis using the same datasets for the purpose of comparison. Table 2 and 3 show the bearing fault diagnosis result obtained by LAMSTAR and CNN for normal speeds at 45 and 30 Hz.

Table 2. LAMSTAR and CNN comparison result on rotating speed at 45 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	100%
Inner race fault	100%	90%
Outer race fault	100%	90%
Cage fault	100%	90%
Ball fault	90%	90%
Over all	98%	92%

Table 3. LAMSTAR and CNN comparison result on rotating speed at 30 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	100%
Inner race fault	100%	80%
Outer race fault	100%	90%
Cage fault	100%	90%
Ball fault	90%	90%
Over all	98%	90%

From Table 2 and Table 3, it can be observed that under the normal speeds, LAMSTAR gives more accurate diagnosis performance than CNN. In addition, the time used to train a LAMSTAR model was 10 times less than CNN. As the shaft speed reduces from 45 Hz to 30 Hz, the diagnostic performance of LAMSTAR remains the same while the diagnostic performance of CNN gets worse.

As pointed out in Van Hecke *et al.* (2014), it is normally difficult to diagnose the bearing faults at low speed in the range of 0.5 Hz and 10 Hz. Using a spectral averaging based approach, they only showed significant fault diagnosis results for rotation speed over 30 Hz. Table 4 to Table 8 show the diagnosis result obtained by LAMSTAR and CNN at relatively low speeds. As shown in the tables, LAMSTAR shows more steady performance than CNN on fault diagnosis with decreasing rotating speed. The classification accuracy from LAMSTAR application decreases from 98% to 96%, while the one from CNN drops from 88% to 80%. The results presented in this paper show the powerful diagnostic performance of deep learning based approach for even relatively low speeds.

Table 4. LAMSTAR and CNN comparison result on rotating speed at 10 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	100%
Inner race fault	100%	90%
Outer race fault	100%	90%
Cage fault	100%	80%
Ball fault	90%	80%
Over all	98%	88%

Table 5. LAMSTAR and CNN comparison result on rotating speed at 8 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	100%
Inner race fault	100%	90%
Outer race fault	100%	90%
Cage fault	100%	80%
Ball fault	90%	80%
Over all	98%	88%

Table 6. LAMSTAR and CNN comparison result on rotating speed at 6 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	100%
Inner race fault	90%	80%
Outer race fault	100%	80%
Cage fault	100%	80%
Ball fault	90%	80%
Over all	96%	84%

Table 7. LAMSTAR and CNN comparison result on rotating speed at 4 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	100%
Inner race fault	90%	80%
Outer race fault	100%	80%
Cage fault	100%	80%
Ball fault	90%	80%
Over all	96%	84%

Table 8. LAMSTAR and CNN comparison result on rotating speed at 2 Hz

Bearing condition	LAMSTAR accuracy (%)	CNN accuracy (%)
Healthy	100%	80%
Inner race fault	100%	80%
Outer race fault	90%	80%
Cage fault	100%	80%
Ball fault	90%	80%
Over all	96%	80%

5. CONCLUSIONS

In this paper, a deep learning based approach for bearing fault diagnosis using acoustic emission (AE) sensors with big data was presented. The presented approach pre-processes AE signals using short time Fourier transform (STFT) and extracts features. Based on the simply processed AE features, the optimized deep learning structure, LAMSTAR was used to perform bearing fault diagnosis. The presented method was validated with AE data collected from seeded bearing fault tests performed on a bearing test rig in the laboratory. The bearing fault diagnosis performance of the LAMSTAR was also compared with another deep learning method CNN. The results have shown that the LAMSTAR based method gives better performance at both the normal and relative low input shaft speeds.

REFERENCES

Girado, I. J., Sandin, J. D., DeFanti, A. T., & Wolf, K. L. (2003). Real-time camera-based face detection using a modified LAMSTAR neural network system. *Proceedings of SPIE 5015. Application of Artificial Neural Networks in Image Processing*, vol. 8, no. 36, pp. 36-46. March 27, Santa Clara, CA. doi:10.1117/12.477405.

- Girado, J. I., (2004). *Real-time 3d head position tracker system with stereo cameras using a face recognition neural network*, Doctoral dissertation. University of Illinois at Chicago, Chicago, U.S. https://www.evl.uic.edu/documents/giradophdthesis_10_29_04.pdf.
- Graupe, D. (2007). *Principles of Artificial Neural Networks, Second Edition*. Chicago, IL, U.S. World Scientific Publishing Co.
- Graupe, D. & Kordylewski, H. (1997). A large scale memory (LAMSTAR) neural network for medical diagnosis. *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual Conference of the IEEE*, vol. 3, pp. 1332-1335. October 30-November 02, 1997. Chicago, IL. doi: [10.1109/IEMBS.1997.756622](https://doi.org/10.1109/IEMBS.1997.756622).
- He, D., Li, R., Zhu, J., & Zade, M. (2011). Data Mining Based Full Ceramic Bearing Fault Diagnostic System Using AE Sensors. *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2022-2031. doi:10.1109/TNN.2011.2169087.
- He, Y. & Zhang, X. (2012). Approximate entropy analysis of the acoustic emission from defects in rolling element bearings. *Journal of Vibration and Acoustics*, vol. 134, no. 6, pp. 061012. doi:10.1115/1.4007240.
- Hemmati, F., Orfali, W., & Gadala, M. S. (2016). Roller bearing acoustic signature extraction by wavelet packet transform, applications in fault detection and size estimation. *Applied Acoustics*, vol. 104, pp. 101-118. doi:10.1016/j.apacoust.2015.11.003.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, vol. 313, no. 5786, pp. 504-507. doi:10.1126/science.1127647.
- Huang, R., Xi, L., Li, X., Liu, C. R., Qiu, H., & Lee, J. (2007). Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 193-207. doi:10.1016/j.ymsp.2005.11.008.
- Isola, R., Carvallo, R., & Tripathy, A. K. (2012). Knowledge discovery in medical systems using differential diagnosis, LAMSTAR, and k-NN. *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1287-1294. doi:10.1109/TITB.2012.2215044.
- Lei, Y., Lin, J., He, Z., & Kong, D. (2012). A method based on multi-sensor data fusion for fault detection of planetary gearboxes. *Sensors*, vol. 12, no. 2, pp. 2005-2017. doi:10.3390/s120202005.
- Mba, D. (2008). The use of acoustic emission for estimation of bearing defect size. *Journal of Failure Analysis and Prevention*, vol. 8, no. 2, pp. 188-192. doi:10.1007/s11668-008-9119-8.

- Morhain, A. & Mba, D. (2003). Bearing defect diagnosis and acoustic emission, *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 217, no. 4, pp. 257 – 272. doi: 10.1243/135065003768618614.
- Nienhaus, K., Boos, F.D., Garate, K., & Baltes, R. (2012). Development of acoustic emission (AE) based defect parameters for slow rotating roller bearings, *Journal of Physics: Conference Series*, vol. 364, no. 1, June 18–20, Huddersfield, UK.
- Paya, B. A., Esat, I. I., & Badi, M. N. M. (1997). Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical systems and signal processing*, vol. 11, no. 5, pp. 751-765. doi:10.1006/mssp.1997.0090.
- Samantha, B. (2004). Artificial neural networks and genetic algorithms for gear fault detection. *Mechanical Systems and Signal Processing*, vol. 18, pp. 1273-1282. doi:10.1016/S0888-3270(03)00020-7.
- Samanta, B. & Al-Balushi, K.R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical systems and signal processing*, vol. 17, no. 2, pp.317-328. doi:10.1006/mssp.2001.1462.
- Samantha, B. & Nataraj, C (2009). Use of particle swarm optimization for machinery fault detection. *Engineering Application and Artificial Intelligence*, vol. 22, pp. 308-316. doi:10.1016/j.engappai.2008.07.006.
- Sivaramakrishna, A., & Graupe, D. (2004). Brain tumor demarcation by applying a LAMSTAR neural network to spectroscopy data. *Neurological Research: A Journal of Progress in Neurosurgery, Neurology and Neuro Sciences*, vol. 26, no. 6, pp. 613-621. doi: 10.1179/016164104225017802.
- Wang, D., & Tse, W. P. (2015). Prognostics of slurry pumps based on a moving-average wear degradation index and a general sequential Monte Carlo method. *Mechanical Systems and Signal Processing*, vol. 56, pp. 213-229. doi:10.1016/j.ymsp.2014.10.010.
- Waxman, J. A., Graupe, D., & Carley, D. W. (2010). Automated prediction of apnea and hypopnea, using a LAMSTAR artificial neural network. *American Journal of Respiratory and Critical Care Medicine*, vol. 181, no. 7, pp. 727-733. doi: 10.1164/rccm.200907-1146OC
- Yoon, J.M., He, D. & Qiu, B. (2013). Full ceramic bearing fault diagnosis using LAMSTAR neural network. *In Prognostics and Health Management (PHM), IEEE Conference on*, pp. 1-9. June 24-27, Gaithersburg, MD. doi: 10.1109/ICPHM.2013.6621427.
- Van Hecke, B., He, D., & Qu, Y. (2014). On the Use of Spectral Averaging of Acoustic Emission Signals for Bearing Fault Diagnostics, *ASME Journal of Vibration and Acoustics*, vol. 136, no. 6, pp. 1-13. doi: 10.1115/1.4028322.

BIOGRAPHIES



Miao He received her B.E. degree in Vehicle Engineering from University of Science and Technology Beijing, Beijing, China. She then received M.S. degree in Mechanical Engineering from Purdue University, Calumet, IN. She worked as research assistant and lab assistant during period of proceeding of M.S. degree. She is currently a Ph.D. student in the Department of Mechanical and Industrial Engineering at University of Illinois at Chicago. Her research interests include: digital signal processing, machinery health monitoring and fault diagnosis, prognostics, artificial neural network applications.



David He received his B.S. degree in Metallurgical Engineering from Shanghai University of Technology, China, MBA degree from the University of Northern Iowa, and Ph.D. degree in Industrial Engineering from the University of Iowa in 1994. Dr. He is a Professor and Director of the Intelligent Systems Modeling & Development Laboratory in the Department of Mechanical and Industrial Engineering at The University of Illinois-Chicago. Dr. He's research areas include: machinery health monitoring, diagnosis and prognosis, complex systems failure analysis, quality and reliability engineering, and manufacturing systems design, modeling, scheduling and planning.



Eric Bechhoefer received his B.S. in Biology from the University of Michigan, his M.S. in Operations Research from the Naval Postgraduate School, and a Ph.D. in General Engineering from Kennedy Western University. He is a former Naval Aviator who has worked extensively on condition based maintenance, rotor track and balance, vibration analysis of rotating machinery and fault detection in electronic systems. Dr. Bechhoefer is a board member of the Prognostics Health Management Society, and a member of the IEEE Reliability Society.